

# ChatGPT and Software Testing Education: Promises & Perils

Year: 2023 | Citations: 202 | Authors: Sajed Jalil, Suzzana Rafi, Thomas D. Latoza, Kevin Moran, Wing Lam

---

## Abstract

Over the past decade, predictive language modeling for code has proven to be a valuable tool for enabling new forms of automation for developers. More recently, we have seen the advent of general purpose "large language models", based on neural transformer architectures, that have been trained on massive datasets of human written text, which includes code and natural language. However, despite the demonstrated representational power of such models, interacting with them has historically been constrained to specific task settings, limiting their general applicability. Many of these limitations were recently overcome with the introduction of ChatGPT, a language model created by OpenAI and trained to operate as a conversational agent, enabling it to answer questions and respond to a wide variety of commands from end users. The introduction of models, such as ChatGPT, has already spurred fervent discussion from educators, ranging from fear that students could use these AI tools to circumvent learning, to excitement about the new types of learning opportunities that they might unlock. However, given the nascent nature of these tools, we currently lack fundamental knowledge related to how well they perform in different educational settings, and the potential promise (or danger) that they might pose to traditional forms of instruction. As such, in this paper, we examine how well ChatGPT performs when tasked with answering common questions in a popular software testing curriculum. We found that given its current capabilities, ChatGPT is able to respond to 77.5% of the questions we examined and that, of these questions, it is able to provide correct or partially correct answers in 55.6% of cases, provide correct or partially correct explanations of answers in 53.0% of cases, and that prompting the tool in a shared question context leads to a marginally higher rate of correct answers and explanations. Based on these findings, we discuss the potential promises and perils related to the use of ChatGPT by students and instructors.